Analytic Variations on Redundancy Rates of Renewal Processes

Philippe Flajolet and Wojciech Szpankowski, *Senior Member, IEEE*

***Abstract*—Csiszár and Shields have proved that the minimax redundancy for a class of (stationary) renewal processes is $\Theta(\sqrt{n})$ where n is the block length. This interesting result provides a nontrivial bound on redundancy for a nonparametric family of processes. The present paper gives a precise estimate of the redundancy rate for such (nonstationary) renewal sources, namely,**

$$\frac{2}{\log 2}\sqrt{\left(\frac{\pi^2}{6}-1\right)n}+O(\log n).$$

This asymptotic expansion is derived by complex-analytic methods that include generating function representations, Mellin transforms, singularity analysis. and saddle-point estimates. This work places itself within the framework of analytic information theory.

***Index Terms*—Analytic information theory, Mellin transform, partitions of integers, redundancy, renewal processes, saddle-point method, tree function, universal coding.**

I. Introduction

Recent years have seen a resurgence of interest in redundancy rates of lossless coding; see [4], [15], [17], [19]–[22], [26], [27], [29]. The redundancy-rate problem of universal fixed-to-variable length coding for a class of sources consists in determining by how much the actual code length exceeds the optimal (ideal) code length. In a minimax scenario, one finds the best code for the worst source. While Shields [20] proved that there is no function $o(n)$ which is a rate bound on the redundancy for the class of *all ergodic* processes, it has been known for some time (cf. [17], [27]) that, for certain parametric families of sources (e.g., memoryless and Markov sources), the redundancy can be as small as $\Theta(\log n)$ where n is the block length. There was no interesting bound for a class of sources that lies between $\Theta(\log n)$ and general $o(n)$ until recently, when Csiszár and Shields [4] designed a renewal class of sources that yields a $\Theta(\sqrt{n})$ bound. Still, one would like to know more about this bound. What is, for instance, the constant in front of $\sqrt{n}$, if there is one? (See [15] for an example where fluctuations are involved in the redundancy rate.) Is the next term of the redundancy $100\sqrt{n/\log n}$ or $2\log n$? And so forth. In this paper, we address these questions by providing an asymptotic expansion of the redundancy for renewal sources up to the $O(\log n)$ term.

Regarding coding theory, we shall follow the notation and the presentation from [4]. A code $C_n\colon \mathcal{A}^n \to \{0,1\}^*$ is defined as an injective mapping from the set $\mathcal{A}^n$ of all sequences of length n over the finite alphabet $\mathcal{A}$ to the set $\{0,1\}^*$ of all binary sequences. We consider here only uniquely decodable (lossless) coding. A message of arbitrary length n with letters indexed from 1 to n is denoted by x_1^n, so that $x_1^n \in \mathcal{A}^n$. We write X_1^n to denote the random variable representing a message of length n. Given a probabilistic source model, we let $P(x_1^n)$ be the probability of the message x_1^n; given a code C_n, we let $L(C_n, x_1^n)$ be the code length for x_1^n. Information–theoretic quantities are expressed in binary logarithms written $\lg := \log_2$. We also write $\log := \ln$.

From Shannon's works, we know that the entropy

$$H_n(P) = -\sum_{x_1^n} P(x_1^n)\lg P(x_1^n)$$

is an absolute lower bound on the expected code length, and $-\lg P(x_1^n)$ can be viewed as the "ideal" code length. The next natural question is to ask by how much the length $L(C_n, x_1^n)$ of the code C_n differs from the ideal code length, either for individual sequences or on average. Thus, the *pointwise redundancy* $R_n(C_n, P; x_1^n)$ and the *average redundancy* $\overline{R}_n(C_n, P)$ are defined as

$$\begin{aligned} R_n(C_n, P; x_1^n) &= L(C_n, x_1^n) + \lg P(x_1^n)\\ \overline{R}_n(C_n, P) &= \mathbf{E}_P[R_n(C_n, P; X_1^n)]\\ &= \mathbf{E}[L(C_n, X_1^n)] - H_n(P) \end{aligned}$$

where the underlying probability measure P represents a particular source model and $\mathbf{E}$ denotes the expectation. Another natural measure of code performance is the *maximal* redundancy defined as

$$R_n^*(C_n, P) = \max_{x_1^n}\{R_n(C_n, P; x_1^n)\}.$$

Observe that while the pointwise redundancy can be negative, maximal and average redundancies cannot, by Kraft's inequality and Shannon's source-coding theorem, respectively.

In practice, the source probabilities are unknown, hence the desire to design codes for a whole class of source models $\mathcal{S}$. When the source is known, the redundancy can be as low as 1 bit, as demonstrated, for example, by Shannon codes. Therefore, for unknown probabilities, the redundancy rate can be also

Manuscript received January 28, 1999; revised May 2, 2002. The work of P. Flajolet was supported in part by the IST Programme of the EU under Contract IST-1999-14186 (ALCOM-FT). The work of W. Szpankowski was supported by the National Science Foundation (NSF) under Grants CCR-9804760 and CCR-0208709, and by sponsors of CERIAS at Purdue University under Contract 1419991431A. The material in this paper was presented in part at the IEEE International Symposium on Information Theory, Sorrento, Italy, June 2000.

P. Flajolet is with the Algorithms Project, INRIA-Rocquencourt, F78153 Le Chesnay, France (e-mail: Philippe.Flajolet@inria.fr).

W. Szpankowski is with the Department of Computer Science, Purdue University, West Lafayette, IN 47907 USA (e-mail: spa@cs.purdue.edu).

Communicated by I. Csiszár, Associate Editor for Shannon Theory.

Digital Object Identifier 10.1109/TIT.2002.804115.

0018-9448/02$17.00

viewed as the penalty paid for estimating the underlying probability measure. The *weak redundancy-rate problem* for the class $\mathcal{S}$ of sources can be roughly viewed as finding a bound on the redundancy rate for a sequence of codes C_n over all $P \in \mathcal{S}$ (cf. [4]). The (asymptotic) *strong redundancy-rate problem* consists in determining for a class $\mathcal{S}$ of source models the rate of growth either of the average minimax redundancy or the maximal minimax redundancy, respectively

$$\overline{R}_n(\mathcal{S}) = \min_{C_n} \max_{P \in \mathcal{S}} \{\overline{R}_n(C_n, P)\}$$
$$R_n^*(\mathcal{S}) = \min_{C_n} \max_{P \in \mathcal{S}} \{R_n^*(C_n, P)\}$$

as $n \to \infty$. In this paper, we deal with the stronger version, namely. with the maximal minimax redundancy R_n^* for arbitrary renewal sources.

The redundancy rate problem is typical of a situation where second-order asymptotics play a crucial role since the leading term of $L(C_n, X_1^n)$ is known to be nH, where H is the entropy rate. This problem is an ideal candidate for "analytic information theory" that applies complex-analytic tools to information theory. As argued by Odlyzko [16]: "*Analytic methods are extremely powerful and when they apply, they often yield estimates of unparalleled precision*." We shall see this principle at work for the redundancy problem. (Other examples are provided by [12], [13], [15], [22], [24], [26].) In fact, in his *1997 Shannon Lecture* [30], Ziv presented compelling arguments for "backing off" to a certain degree from the (first-order) asymptotic analysis of information systems in order to predict the behavior of real systems where we always face *finite* (and often small) lengths (of sequences, files, codes, etc.) One way of overcoming these difficulties is to increase the accuracy of asymptotic analysis and replace first-order analyses (e.g., a leading term of the average code length) by more complete asymptotic expansions, thereby extending their range of applicability to smaller values while providing more accurate analyses (like constructive error bounds, large deviations, local or central limit laws).

A substantial literature is available on the redundancy problem. Some known results are listed in what follows.

- If $\mathcal{M}$ is a class of independent and identically distributed (i.i.d.) processes or a class of Markov chains, or more generally, the process belongs to a finitely parameterizable class of dimension K, then Rissanen [17] (cf. also [26]) established

$$\overline{R}_n(\mathcal{M}) \sim R_n^*(\mathcal{M}) \sim \frac{K}{2} \lg n.$$

It was also found in [26], [27] that the next term of $\overline{R}_n(\mathcal{M})$ and $R_n^*(\mathcal{M})$ is $O(1)$. In fact, Szpankowski [22] (cf. also [26]) has established a full asymptotic expansion for $R_n^*(\mathcal{M})$ for memoryless sources over an m-ary alphabet, namely

$$R_n^*(\mathcal{M}) = \frac{m-1}{2} \lg\left(\frac{n}{2}\right) + \lg\left(\frac{\sqrt{\pi}}{\Gamma(m/2)}\right) + \cdots$$

where $\Gamma(x)$ is the Euler gamma function. In passing, we observe that when the alphabet size m is large, the second-order terms may contribute significantly to R_n^*. More importantly, the above formula is true only when m is fixed, while in some applications m may depend on n (e.g., image size is comparable to image alphabet). Then, one needs a *uniform* asymptotic expansion of R_n^*, and, clearly, second-order terms will *contribute* to the final outcome.

- Csiszár and Shields [4] have studied order-r stationary Markov renewal sequences in which a 1 is inserted every $T_0, T_1, \ldots$ of 0's, where $\{T_i\}$ is either an i.i.d. or Markov, or r-order Markov process. We denote such sources as $\mathcal{R}_r$. The authors of [4] proved that

$$\overline{R}_n(\mathcal{R}_r) = R_n^*(\mathcal{R}_r) = \Theta(n^{(r+1)/(r+2)}), \qquad \text{for } r = 1, 2, \ldots$$

which specializes to $\Theta(\sqrt{n})$ when $r = 0$.

- Shields [20] proved that there is no function $\rho(n) = o(n)$ which is a weak-rate bound for the class of all ergodic processes.
- Louchard and Szpankowski [15], Savari [19], Wyner [29], and Jacquet and Szpankowski [14] proved that the Lempel–Ziv codes in the class of i.i.d. and Markov processes have either rate $\Theta(n/\log n)$ (for LZ'78) or $\Theta(n \log\log n/\log n)$ (for LZ'77 code). Interestingly, in [15] it was shown that for LZ'78 the bound $\Theta(n/\log n)$ cannot be improved to an asymptotic equivalence since a fluctuating function is involved. More precisely, for a binary alphabet with 0's generated with probability p and 1's with probability $q = 1 - p$, the authors of [15] showed that

$$\overline{R}_n(LZ) = H(\chi + \delta(n)) \frac{n}{\log n} + O\left(\frac{n \log\log n}{\log^2 n}\right)$$

where $H = -p \lg p - (1-p)\lg(1-p)$ is the entropy rate and χ is an explicitly determined constant. What is more surprising is the occurrence of the function $\delta(x)$ that fluctuates with mean zero and a tiny amplitude for $\log p/\log q$ rational (the amplitude of $\delta(x)$ is smaller than 10^{-6} for the unbiased case, where $p = q = 0.5$), but satisfies $\lim_{x\to\infty} \delta(x) = 0$ otherwise.

In this paper, we revisit Csiszár and Shields renewal process and present a precise analysis of the maximal minimax redundancy rate for the class of basic renewal source $\mathcal{R}_0$ corresponding to $r = 0$. However, instead of analyzing stationary renewal sequence we consider a nonstationary renewal sequence that starts with a 1 (see Section II for details). Our main result is as follows.

Theorem 1: Consider the class $\mathcal{R}_0$ of nonstationary renewal sources. Let $c = \frac{\pi^2}{6} - 1 \approx 0.645$. Then the minimax redundancy $R_n^*(\mathcal{R}_0)$ satisfies

$$\frac{2}{\log 2}\sqrt{cn} - \frac{5}{8}\lg n + \frac{1}{2}\log\log n \le R_n^*(\mathcal{R}_0) \le \frac{2}{\log 2}\sqrt{cn} - \frac{1}{8}\lg n + \frac{1}{2}\log\log n \quad (1)$$

for large n.

We believe that a more refined analysis of Lemma 1 that follows could allow us to conclude that the upper bound of (1) is the correct asymptotic expression for R_n^* up to $O(1)$ term. Actually, some recent results of [6] may lead to a precise calculation of the $O(1)$ term as well. Furthermore, it would be interesting to see if the constant of the leading term of the maximal minimax redundancy $R_n^*(\mathcal{R}_0)$ is the same as for the average minimax redundancy $\overline{R}_n(\mathcal{R}_0)$. The answer to the latter question would shed some light on a wider problem, namely, whether $\overline{R}_n \sim R_n^*$ for a class of stationary and ergodic processes. Some preliminary results in this direction can be found in [6].

This paper is organized as follows. In Section II, the problem is reduced to estimating a certain combinatorial sum that is of independent interest (cf. (7)). Next, we present the main ingredients of the proof in Section III. The heart of the analysis is Lemma 2 that is established in Section IV. The proof is analytic and uses such diverse tools as the Mellin transform, singularity analysis, and the saddle point method. We believe that the analytic approach discussed in this section is of general interest and hope it may find further applications in information theory. For this reason, as well as for convenience of exposition, we adopt a tabular presentation of the two main tools used here: the Mellin transform (Fig. 1) and the saddle point method (Fig. 2).

II. Reduction to a Combinatorial Sum

We start with a precise definition of the class $\mathcal{R}_0$ of nonstationary renewal process and its associated sources. Let $T_1, T_2 \cdots$ be a sequence of i.i.d. positive-valued random variables with common distribution $Q(j) = \Pr\{T_1 = j\}$. Throughout we assume that $\mathbf{E}[T_1] < \infty$. The process $T_1, T_1 + T_2, \ldots$ is called the (nonstationary) renewal process. With such a renewal process there is associated a *binary renewal sequence* that is a $0,$ 1-sequence in which the 1's occur exactly at the renewal epochs $T_1,\ T_1{+}T_2$, etc. Accordingly, we start the renewal sequence with a 1 put at the zeroth position. In passing we observe that since P and Q determine one another, we freely identify the underlying probability measure P defined on $\{0,\,1\}^\infty$ with the distribution Q that it induces.

We should mention that Csiszár and Shields considered a *stationary* renewal sequence so that starts with some initial 0's of length T_0 satisfying

$$\Pr\{T_0 = i\} = \mathbf{E}[T_1]^{-1} \sum_{j \ge i} Q(j).$$

We now briefly discuss the mathematical aspect of the redundancy problem. Our goal is to derive asymptotics of

$$R_n^*(\mathcal{S}) = \min_{C_n} \sup_{Q \in \mathcal{S}} \max_{x_1^n} \left[L(C_n,\, x_1^n) + \lg P(x_1^n)\right]$$

where the supremum is taken over all corresponding distributions Q or P. Shtarkov's maximum-likelihood technique [21] implies

$$\lg\left(\sum_{x_1^n} \sup_Q P(x_1^n)\right) \le R_n^*(\mathcal{S}) \le \lg\left(\sum_{x_1^n} \sup_Q P(x_1^n)\right) + 1. \tag{2}$$

Indeed, for the lower bound, Shtarkov [21] observed that

$$q(x_1^n) := \frac{\sup_Q P(x_1^n)}{\sum_{x_1^n} \sup_Q P(x_1^n)}$$

is a probability distribution, and by Kraft's inequality there exists $\tilde{x}_1^n$ such that

$$L(C_n,\, \tilde{x}_1^n) \ge -\lg q(\tilde{x}_1^n).$$

This implies the lower bound as follows:

$$\begin{aligned} R_n^*(\mathcal{S}) &\ge \min_{C_n} \sup_Q \left[\lg(1/q(\tilde{x}_1^n)) + \lg P(\tilde{x}_1^n)\right] \\ &\ge \lg\left(\sum_{x_1^n} \sup_Q P(x_1^n)\right) \\ &\quad + \min_{C_n} \sup_Q \left[\lg P(\tilde{x}_1^n) - \lg \sup_Q P(\tilde{x}_1^n)\right] \\ &= \lg\left(\sum_{x_1^n} \sup_Q P(x_1^n)\right). \end{aligned}$$

For the upper bound, Shtarkov proposed to take the Shannon code $\tilde{C}_n$ for the distribution $q(x_1^n)$ of length

$$L(\tilde{C}_n,\, x_1^n)) = \lceil -\lg q(x_1^n) \rceil$$

which gives the desired upper bound by the following simple implications:

$$\begin{aligned} R_n^*(\mathcal{S}) &\le \sup_Q \max_{x_1^n} \left[\lceil -\lg 1/q(x_1^n) \rceil + \lg P(x_1^n)\right] \\ &\le \lg\left(\sum_{x_1^n} \sup_Q P(x_1^n)\right) + 1. \end{aligned}$$

This completes the derivation of (2).

We shall derive asymptotics of $R_n^*(\mathcal{S})$ up to $O(\log n)$ term. We first define

$$r_n^* = \sum_{x_1^n} \sup_Q P(x_1^n). \tag{3}$$

By Shtarkov's inequality (2), we know that $\lg r_n^*$ is within $O(1)$ from the maximal redundancy $R_n^*(\mathcal{R}_0)$.

In order to estimate r_n^* we need to compute $P(x_1^n)$ and analyze $\sup_Q P(x_1^n)$ that we discuss next. Observe that the nonstationary renewal sequence $x_0^n = 1x_1^n$ can be represented as

$$x_0^n = 10^{\alpha_1}1 \cdots 10^{\alpha_l}1\underbrace{0 \cdots 0}_{k^*}$$

where $0 \le \alpha_i \le n$ for $i = 1,\, \ldots,\, l \le n$. Let k_m be the number of i such that $\alpha_i = m$, where $m = 0,\, 1,\, \ldots,\, n-1$. Then

$$P(x_1^n) = Q^{k_0}(1) Q^{k_1}(2) \cdots Q^{k_{n-1}}(n) \Pr\{T_1 > k^*\} \tag{4}$$

subject to $Q(1) + Q(2) + \cdots + Q(n) \le 1$ and

$$k_0 + 2k_1 + \cdots + nk_{n-1} + k^* = n. \tag{5}$$

To simplify somewhat the next presentation we set $q_i = Q(i+1)$, that is, q_i describes the distribution of the number of zeros between two renewals. (By definition, we set $q_{-1} = Q(0) = 0$.)

As a check, we verify that $P(x_0^n)$ is a probability distribution, that is, $\sum_{x_0^n} P(x_0^n) = 1$ for all $n \ge 0$, where, by definition, $P(x_0^0 = 1) = 1$. We prove it using generating functions (cf. also Section IV and [9], [23]). In particular, we shall show that for $|z| \le 1$, we have

$$\sum_{n=0}^{\infty} z^n \sum_{x_0^n} P(x_0^n) = \frac{1}{1-z}. \tag{6}$$

Indeed, let $k = k_0 + \cdots k_{n-1}$ where $k_0, \ldots, k_{n-1}$ satisfy (5). Introducing two new variables u and v, recalling that

$$q_i = \Pr\{T_1 = i+1\} = Q(i+1)$$

and defining $F(v) = \sum_{i\ge 0} q_i v^i$, we have with $n^* = n - k^*$ (the first three lines of the below summation is over tuples $(k_0, \ldots, k_{n-1})$ satisfying (5) and $k = k_0 + \cdots k_{n-1}$, but, for simplicity, we are not showing it below explicitly)

$$\begin{aligned}\sum_{n\ge 0}\sum_{k\ge 0}\sum_{k^*\ge 0} &\binom{k}{k_0,\, k_1,\, \ldots} q_0^{k_0} q_1^{k_1}\cdots \\ &(1-q_0-\cdots-q_{k^*-1}) z^n u^k v^{k^*}\\ &= \sum_{n^*\ge 0}\sum_{k\ge 0}\binom{k}{k_0,\, k_1,\, \ldots} q_0^{k_0}q_1^{k_1}\cdots z^{n^*} u^k \frac{1-zvF(zv)}{1-zv}\\ &= \frac{1-zvF(zv)}{1-zv}\sum_{k\ge 0} u^k \sum_{n^*\ge 0}\binom{k}{k_0,\, k_1,\, \ldots}\\ &\quad \cdot (q_0 z)^{k_0}\left(q_1 z^2\right)^{k_1}\cdots\\ &= \frac{1-zvF(zv)}{1-zv}\sum_{k\ge 0}(zuF(z))^k\\ &= \frac{1-zvF(zv)}{1-zv}\,\frac{1}{1-uzF(z)}.\end{aligned}$$

Substituting in the last line $u = 1$ and $v = 1$ proves (6).

It turns out that the problem of estimating r_n^* can be reduced to the evaluation of a purely combinatorial sum r_m defined in (7) that we study throughout the rest of the paper (by convention here $0^0 = 1$)

$$\begin{cases} r_m = \displaystyle\sum_{k=0}^{m} r_{m,k}\\ r_{m,k} = \displaystyle\sum_{\mathcal{P}(m,k)}\binom{k}{k_0\cdots k_{m-1}}\left(\frac{k_0}{k}\right)^{k_0}\cdots\left(\frac{k_{m-1}}{k}\right)^{k_{m-1}}\end{cases} \tag{7}$$

where $\mathcal{P}(m, k)$ denotes the set of partitions of m into k summands, that is, the collection of tuples of nonnegative integers $(k_0, k_1, \ldots, k_{m-1})$ satisfying

$$m = k_0 + 2k_1 + \cdots + mk_{m-1} \tag{8}$$

$$k = k_0 + k_1 + \cdots + k_{m-1}. \tag{9}$$

It can also be observed that the quantity r_m has an intrinsic meaning by its own. Let $\mathcal{W}_m$ denote the set of all m^m sequences of length m over the alphabet $\{0, \ldots, m-1\}$. For a sequence w, take k_j to be the number of letters j in w. Then each sequence w carries a "maximum-likelihood probability" $\pi_{\rm ML}(w)$ (given later by (14)): this is the probability that w gets assigned in the Bernoulli model that makes it most likely. The quantity r_m is also

$$r_m = \sum_{w\in\mathcal{W}_m}\pi_{\rm ML}(w).$$

Returning to our problem, we present in the following lower and upper bounds for r_n^* in terms of r_n.

Lemma 1: For all $n \ge 0$

$$r_{n+1} - 1 \le r_n^* \le \sum_{m=0}^{n} r_m \tag{10}$$

where r_n is defined in (7)–(9).

Proof: Let, as before, $q_i = Q(i+1) = \Pr\{T = i+1\}$. By (4), we have

$$P(x_1^n) = q_0^{k_0} q_1^{k_1}\cdots q_{n-1}^{k_{n-1}}(1-q_0-q_1-\cdots-q_{k^*-1}) \tag{11}$$

subject to $q_0 + \cdots + q_{n-1} \le 1$, $k^* \ge 0$, and

$$k_0 + 2k_1 + \cdots + nk_{n-1} \le n. \tag{12}$$

Observe that in (12), we have $\le n$ instead of $= n$ (cf. (8) where m is replaced by n). We denote such a set of partitions (i.e., satisfying (12) and (9) in which m is replaced by n) as $\mathcal{P}(\le n, k)$.

Upper Bound: We now proceed as follows:

$$\begin{aligned}\sup_q\, & q_0^{k_0}q_1^{k_1}\cdots q_{n-1}^{k_{n-1}}(1-q_0-q_1-\cdots-q_{k^*-1})\\ &\le \sup_q\, q_0^{k_0}q_1^{k_1}\cdots q_{n-1}^{k_{n-1}} &(13)\\ &= \left(\frac{k_0}{k}\right)^{k_0}\cdots\left(\frac{k_{n-1}}{k}\right)^{k_{n-1}} &(14)\end{aligned}$$

provided $\mathcal{P}(\le n, k)$ holds. The second line of the above is a consequence of $(1-q_0-q_1-\cdots-q_{k^*-1}) \le 1$. The last line of the above follows from solving a simple optimization problem with the constraints $q_0 + q_1 + \cdots + q_{n-1} \le 1$. By the Kuhn–Tucker condition or otherwise, it is seen that

$$\frac{q_i}{q_j} = \frac{k_i}{k_j}, \qquad \text{for } 0 \le i,\, j \le n-1$$

maximizes the product $q_0^{k_0}q_1^{k_1}\cdots q_{n-1}^{k_{n-1}}$ provided n and k satisfy $\mathcal{P}(\le n, k)$. Thus, (14) is established. Since $\mathcal{P}(\le n, k) = \bigcup_{i=0}^{n}\mathcal{P}(n-i, k)$, we finally get the upper bound for r_n^*.

Lower Bound: The lower bound is more intricate. We first observe that the last term of the probability $P(x_1^n)$ can be estimated as

$$(1-q_0-\cdots-q_{k^*-1}) = \Pr\{T > k^*\} \ge \Pr\{T = k^*+1\} = q_{k^*}.$$

In other words, we add an additional 1 at the end of the sequence (making it of length $n+1$), but then the last k^* zeros fall into the same distribution as the previous ones, and can be handled by the same optimization technique as in the upper bound case. A simple calculation reveals that

$$\begin{aligned}\sup_q\, & q_0^{k_0}q_1^{k_1}\cdots q_{k^*}^{k^*+1}\cdots q_{n-1}^{k_{n-1}}\\ &= \left(\frac{k_0}{k+1}\right)^{k_0}\cdots\left(\frac{k_{k^*}+1}{k+1}\right)^{k^*+1}\cdots\left(\frac{k_{n-1}}{k+1}\right)^{k_{n-1}}\end{aligned}$$

subject to $k^* \ge 0$ with n and k satisfying $\mathcal{P}(n+1, k+1)$. We have the following chain of inequalities where below $\mathcal{P}(n, k)_j$

is the set of n-tuples $(k_0, k_1, \ldots, k_{n-1}) \in \mathcal{P}(n, k)$ satisfying $k_j > 0$, and we write $j = k^*$ for simplicity

$$\begin{aligned}
r_n^* &\geq \sum_{k=0}^{n} \sum_{j\geq 0} \sum_{\mathcal{P}(n-j,\,k)} \binom{k}{k_0,\,\ldots} \left(\frac{k_0}{k+1}\right)^{k_0} \cdots \left(\frac{k_j+1}{k+1}\right)^{k_j+1} \cdots \\
&= \sum_{k=0}^{n} \sum_{j\geq 0} \sum_{\mathcal{P}(n-j,\,k)} \frac{k_j+1}{k+1} \binom{k+1}{k_0,\,k_1,\,\ldots,\,k_j+1,\,\ldots} \cdot \left(\frac{k_0}{k+1}\right)^{k_0} \cdots \left(\frac{k_1}{k+1}\right)^{k_1} \cdots \left(\frac{k_j+1}{k+1}\right)^{k_j+1} \cdots \\
&= \sum_{k=0}^{n} \sum_{j\geq 0} \sum_{\mathcal{P}(n+1,\,k+1)_j} \frac{k_j}{k+1} \binom{k+1}{k_0,\,k_1,\,\ldots,\,k_j,\,\ldots} \cdot \left(\frac{k_0}{k+1}\right)^{k_0} \cdots \left(\frac{k_1}{k+1}\right)^{k_1} \cdots \left(\frac{k_j}{k+1}\right)^{k_j} \cdots \\
&\overset{\text{(A)}}{=} \sum_{k=0}^{n} \sum_{j\geq 0} \sum_{\mathcal{P}(n+1,\,k+1)} \frac{k_j}{k+1} \binom{k+1}{k_0,\,\ldots} \left(\frac{k_0}{k+1}\right)^{k_0} \cdot \left(\frac{k_1}{k+1}\right)^{k_1} \cdots \\
&\overset{\text{(B)}}{=} \sum_{k=0}^{n} \sum_{\mathcal{P}(n+1,\,k+1)} \sum_{j\geq 0} \frac{k_j}{k+1} \binom{k+1}{k_0,\,\ldots} \left(\frac{k_0}{k+1}\right)^{k_0} \cdot \left(\frac{k_1}{k+1}\right)^{k_1} \cdots \\
&= \sum_{k=0}^{n} \sum_{\mathcal{P}(n+1,\,k+1)} \binom{k+1}{k_0,\,\ldots} \left(\frac{k_0}{k+1}\right)^{k_0} \left(\frac{k_1}{k+1}\right)^{k_1} \cdots \\
&= r_{n+1} - 1.
\end{aligned}$$

Note that in step (A) we can "add" those terms with $k_j = 0$ because of the factor $k_j/(k+1)$, while in (B), we use the fact that $\sum_j k_j = k+1$. This proves the lower bound and Lemma 1. □

III. A Streamlined Analysis

Our goal is to estimate asymptotically r_n through asymptotics of $r_{n,k}$. A difficulty of finding such asymptotics stems from the factor $k!/k^k$ present in the definition (7) of $r_{n,k}$. We circumvent this problem by analyzing a related pair of sequences, namely, s_n and $s_{n,k}$, that are defined as

$$\begin{cases} s_n = \displaystyle\sum_{k=0}^{n} s_{n,k} \\ s_{n,k} = e^{-k} \displaystyle\sum_{\mathcal{P}(n,\,k)} \frac{k_0^{k_0}}{k_0!} \cdots \frac{k_{n-1}^{k_{n-1}}}{k_{n-1}!}. \end{cases} \tag{15}$$

The translation from s_n to r_n is most conveniently expressed in probabilistic terms. Introduce the random variable K_n whose probability distribution is $s_{n,k}/s_n$, that is,

$$\varpi_n: \qquad \Pr\{K_n = k\} = \frac{s_{n,k}}{s_n} \tag{16}$$

where ϖ_n denotes the distribution. Then Stirling's formula yields

$$\begin{aligned}
\frac{r_n}{s_n} &= \sum_{k=0}^{n} \frac{r_{n,k}}{s_{n,k}} \frac{s_{n,k}}{s_n} = \mathbf{E}\left[(K_n)! K_n^{-K_n} e^{K_n}\right] \\
&= \mathbf{E}\left[\sqrt{2\pi K_n}\right] + O\left(\mathbf{E}\left[K_n^{-\frac{1}{2}}\right]\right).
\end{aligned} \tag{17}$$

Thus, the problem of finding r_n reduces to asymptotic evaluations of s_n, $\mathbf{E}[\sqrt{K_n}]$, and $\mathbf{E}[K_n^{-\frac{1}{2}}]$. The heart of the matter is the following lemma which provides the necessary estimates.

Lemma 2: Let $\mu_n = \mathbf{E}[K_n]$ and $\sigma_n^2 = \mathbf{Var}(K_n)$, where K_n has the distribution ϖ_n defined above in (16). The following holds:

$$s_n \sim \exp\left(2\sqrt{cn} - \frac{7}{8}\log n + d + o(1)\right) \tag{18}$$

$$\mu_n = \frac{1}{4}\sqrt{\frac{n}{c}}\,\log\frac{n}{c} + o\left(\sqrt{n}\right) \tag{19}$$

$$\sigma_n^2 = O(n\log n) = o(\mu_n^2) \tag{20}$$

where $c = \pi^2/6 - 1$, $d = -\log 2 - \frac{3}{8}\log c - \frac{3}{4}\log\pi$.

The somewhat delicate proof of Lemma 2 constitutes the core of the paper and it is deferred till the next section. Once the estimates of Lemma 2 are granted, the moments of order $\pm\frac{1}{2}$ of K_n follow by a standard argument based on concentration of the distribution ϖ_n.

Lemma 3: For large n

$$\mathbf{E}\left[\sqrt{K_n}\right] = \mu_n^{1/2}(1 + o(1)) \tag{21}$$

$$\mathbf{E}\left[K_n^{-\frac{1}{2}}\right] = o(1) \tag{22}$$

where $\mu_n = \mathbf{E}[K_n]$.

Proof: We only prove (21) since (22) is obtained in a similar manner. The upper bound $\mathbf{E}[\sqrt{K_n}] \leq \sqrt{\mathbf{E}[K_n]}$ follows from concavity of the function $\sqrt{x}$. The lower bound follows from concentration of the distribution. Chebyshev's inequality and (20) of Lemma 2 entail, for any arbitrarily small $\varepsilon > 0$

$$\Pr\{|K_n - \mu_n| > \varepsilon\mu_n\} \leq \frac{\mathbf{Var}[K_n]}{\varepsilon^2\mu_n^2} = \frac{\delta(n)}{\varepsilon^2}$$

where $\delta(n) \to 0$ as $n \to \infty$. Then

$$\begin{aligned}
\mathbf{E}\left[\sqrt{K_n}\right] &\geq \sum_{k\geq(1-\varepsilon)\mu_n} \sqrt{k}\,\Pr\{K_n \geq k\} \\
&\geq (1-\varepsilon)^{\frac{1}{2}} \mu_n^{1/2} \Pr\{K_n \geq (1-\varepsilon)\mu_n\} \\
&\geq (1-\varepsilon)^{\frac{1}{2}} \left(1 - \frac{\delta(n)}{\varepsilon^2}\right) \mu_n^{1/2}.
\end{aligned}$$

Hence, for any $\eta > 0$, one has $\mathbf{E}[\sqrt{K_n}] > \mu_n^{1/2}(1-\eta)$ provided n is large enough. This completes the proof. □

In summary, r_n and s_n are related by

$$\begin{aligned}
r_n &= s_n \mathbf{E}\left[\sqrt{2\pi K_n}\right](1 + o(1)) \\
&= s_n\sqrt{2\pi\mu_n}\,(1 + o(1))
\end{aligned}$$

by virtue of (17) and Lemma 3. This leads to

$$R_n := \lg r_n = \lg s_n + \tfrac{1}{2}\lg \mu_n + \lg\sqrt{2\pi} + o(1). \tag{23}$$

At this point, it suffices to apply the estimates provided by Lemmas 2 and 1 to prove the main result. A little calculation needed to prove the upper bound is provided at the end of Section IV (since it is the easiest to do through the generating function machinery introduced in the following section).

IV. Complex Asymptotic Analysis

This section provides precise asymptotic estimates for the quantity s_n and for moments of the distribution ϖ_n as expressed in (19) and (20) of Lemma 2. It turns out that the quantities $s_{n,k}$ and s_n have generating functions $S(z, u)$ and $S(z, 1)$, respectively, that are infinite products involving the tree function of combinatorial analysis. The corresponding coefficient asymptotics are dictated by the behavior at the singularity of greatest weight [7], [23]—in the case at hand, the positive real singularity $z = 1$—so that we start by investigating asymptotics of $S(z, 1)$ as $z \to 1$. This itself requires a dedicated analysis by means of the Mellin transform (cf. [23]). Once the dominant singular behavior of $S(z, 1)$ near $z = 1$ has been found, it remains to extract information on the coefficients s_n. This task involves an appeal to the saddle point method (summarized by Lemma 4) and necessitates some technical concentration condition (Lemma 5). (The whole analysis draws its inspiration from a method of Meinardus in the asymptotic theory of integer partitions; see especially [1, Ch. 6].) Proceeding in this way, the estimate (18) of s_n in Lemma 2 is established. Finally, the method adapts gracefully to moment estimates, yielding the other two estimates (19) and (20) of Lemma 2.

Generating Functions: The expression of $s_{n,k}$ in (15) involves quantities of the form $k^k/k!$. We start by introducing the well-known "tree function" $T(z)$ defined as the solution of

$$T(z) = ze^{T(z)}$$

that is, analytic at 0. The function $T(z)$ satisfies, by the Lagrange inversion theorem

$$T(z) = \sum_{k=1}^{\infty} \frac{k^{k-1}}{k!} z^k \tag{24}$$

for $|z| < e^{-1}$. The tree function owes its name to its rôle in tree enumerations and we refer to the survey paper [3] for algebraic and analytic properties of this important special function of combinatorics.

Next, define the function $\beta(z)$ as

$$\beta(z) = \sum_{k=0}^{\infty} \frac{k^k}{k!} e^{-k} z^k.$$

One has (e.g., by Lagrange inversion again or otherwise)

$$\beta(z) = \frac{1}{1 - T(ze^{-1})}.$$

The quantities s_n and $s_{n,k}$ of (15) have generating functions

$$S_n(u) = \sum_{k=0}^{\infty} s_{n,k} u^k, \qquad S(z, u) = \sum_{n=0}^{\infty} S_n(u) z^n.$$

Then, since (15) involves convolutions of sequences of the form $k^k/k!$, we have

$$\begin{aligned} S(z, u) &= \sum_{\mathcal{P}(n,k)} z^{1k_0 + 2k_1 + \cdots} \left(\frac{u}{e}\right)^{k_0 + \cdots + k_{n-1}} \frac{k_0^{k_0}}{k_0!} \cdots \frac{k_{n-1}^{k_{n-1}}}{k_{n-1}!} \\ &= \prod_{i=1}^{\infty} \beta(z^i u). \end{aligned} \tag{25}$$

We also need access to the first moment $\mu_n = \mathbf{E}[K_n]$ and the second factorial moment $\mathbf{E}[K_n(K_n - 1)]$. They are obtained as

$$\begin{aligned} s_n &= [z^n]S(z, 1) \\ \mu_n &= \frac{[z^n]S'_u(z, 1)}{[z^n]S(z, 1)} \\ \mathbf{E}[K_n(K_n - 1)] &= \frac{[z^n]S''_{uu}(z, 1)}{[z^n]S(z, 1)} \end{aligned}$$

where $[z^n]F(z)$ denotes the coefficient at z^n of $F(z)$, $S'_u(z, 1)$ and $S''_{uu}(z, 1)$ represent the first and the second derivative of $S(z, u)$ at $u = 1$.

Mellin Asymptotics: We deal here with s_n that is accessible through its generating function

$$S(z, 1) = \prod_{i=1}^{\infty} \beta(z^i). \tag{26}$$

The behavior of the generating function $S(z, 1)$ as $z \to 1$ is an essential ingredient of the analysis.

First, the singularity of the tree function $T(z)$ at $z = e^{-1}$ is of the square-root type; see [3]. (This results from the failure of the implicit function theorem at $(z, T) = (e^{-1}, 1)$ where the relation becomes locally quadratic in T.) Hence, near $z = 1$, $\beta(z)$ admits the singular expansion (cf. [3])

$$\beta(z) = \frac{1}{\sqrt{2(1-z)}} + \frac{1}{3} - \frac{\sqrt{2}}{24}\sqrt{(1-z)} + O(1-z).$$

We now turn to the infinite product asymptotics as $z \to 1^-$, with z real. Let $L(z) = \log S(z, 1)$ and $z = e^{-t}$, so that

$$L(e^{-t}) = \sum_{k=1}^{\infty} \log \beta(e^{-kt}). \tag{27}$$

Mellin transform techniques provide an expansion of $L(e^{-t})$ around $t = 0$ (or equivalently $z = 1$) since the sum (27) falls under the *harmonic sum* paradigm of [8], [23]. The Mellin approach is by now a standard technique in the analysis of algorithms. For reader's convenience, we recall its main properties in Fig. 1, following [8], to which we refer globally for detailed validity conditions.

First, the Mellin transform $L^*(s) = \mathcal{M}(L(e^{-t}); s)$ of $L(e^{-t})$ is computed by the harmonic sum property (M3) (see Fig. 1). For $\Re(s) \in (1, \infty)$, the transform evaluates to

$$L^*(s) = \zeta(s)\Lambda(s)$$

where $\zeta(s) = \sum_{n\geq 1} n^{-s}$ is the Riemann zeta function, and

$$\Lambda(s) = \int_0^{\infty} \log \beta(e^{-t}) t^{s-1}\, dt.$$

(M1) Direct and Inverse Mellin Transforms. Let c belong to the *fundamental strip* defined below.

$$f^*(s) := \mathcal{M}(f(x); s) = \int_0^\infty f(x)x^{s-1}dx \qquad \iff \qquad f(x) = \frac{1}{2\pi i}\int_{c-i\infty}^{c+i\infty} f^*(s)x^{-s}ds.$$

(M2) Fundamental Strip. The Mellin transform of $f(x)$ exists in the *fundamental strip* $\Re(s) \in (-\alpha, -\beta)$, where

$$f(x) = O(x^\alpha) \qquad (x \to 0), \qquad\qquad f(x) = O(x^\beta) \qquad (x \to \infty)$$

for $\beta < \alpha$.

(M3) Harmonic Sum Property. By linearity and the scale rule $\mathcal{M}(f(ax); s) = a^{-s}\mathcal{M}(f(x); s)$,

$$f(x) = \sum_{k\geq 0} \lambda_k g(\mu_k x) \qquad \iff \qquad f^*(s) = g^*(s)\sum_{k\geq 0}\lambda_k \mu_k^{-s}.$$

(M4) Mapping Properties (Asymptotic expansion of $f(x)$ and singularities of $f^*(s)$).

$$f(x) = \sum_{(\xi,k)\in A} c_{\xi,k}x^\xi(\log x)^k + O(x^M) \qquad \iff \qquad f^*(s) \asymp \sum_{(\xi,k)\in A} c_{\xi,k}\frac{(-1)^k k!}{(s+\xi)^{k+1}}. \qquad (A.1)$$

— (i) *Direct Mapping.* Assume that $f(x)$ admits as $x \to 0^+$ the asymptotic expansion (A.1) for some $-M < -\alpha$ and $k > 0$. Then for $\Re(s) \in (-M, -\beta)$, the transform $f^*(s)$ satisfies the singular expansion (A.1)

— (ii) *Converse Mapping.* Assume that $f^*(s) = O(|s|^{-r})$ with $r > 1$, as $|s| \to \infty$ and that $f^*(s)$ admits the singular expansion (A.1) for $\Re(s) \in (-M, -\beta)$. Then $f(x)$ satisfies the asymptotic expansion (A.1) at $x = 0^+$.

Fig. 1. Main properties of the Mellin transform.

The subsequent treatment is typical of the Mellin analysis of harmonic sums: the singularity structure of $\Lambda(s)$ is deduced from the asymptotic properties of $\beta(z)$. This gives, in turn, the singularity structure of $L^*(s)$ that is then converted back into an asymptotic expansion of $L(e^{-t})$. In effect, by the direct mapping property (M4), the expansion of $\beta(z)$ at $z = 1$ implies

$$\log\beta(e^{-t}) = -\tfrac{1}{2}\log t - \tfrac{1}{2}\log 2 + O\left(\sqrt{t}\right)$$

so that, collecting local expansions

$$\Lambda(s) \asymp (\Lambda(1))_{s=1} + \left(\frac{1}{2}\frac{1}{s^2} - \frac{1}{2}\frac{\log 2}{s}\right)_{s=0}.$$

On the other hand, classical expansions give (cf. [25])

$$\zeta(s) \asymp \left(\frac{1}{s-1} + \gamma\right)_{s=1} + \left(-\frac{1}{2} - s\log\sqrt{2\pi}\right)_{s=0}.$$

Term-wise multiplication then provides the singular expansion of $L^*(s)$

$$L^*(s) \asymp \left(\frac{\Lambda(1)}{s-1}\right)_{s=1} + \left(-\frac{1}{4s^2} - \frac{\log\pi}{4s}\right)_{s=0}.$$

An application of the converse mapping property (M4) allows us to come back to the original function

$$L(e^{-t}) = \frac{\Lambda(1)}{t} + \frac{1}{4}\log t - \frac{1}{4}\log\pi + O\left(\sqrt{t}\right) \qquad (28)$$

which translates, using $1 - t + O(t^2) = e^{-t} = z$, into

$$L(z) = \frac{\Lambda(1)}{1-z} + \frac{1}{4}\log(1-z) - \frac{1}{4}\log\pi - \frac{1}{2}\Lambda(1) + O\left(\sqrt{1-z}\right). \qquad (29)$$

This computation is finally completed by the evaluation of $c := \Lambda(1)$

$$\begin{aligned} c = \Lambda(1) &= -\int_0^1 \log(1 - T(x/e))\,\frac{dx}{x} \\ &= -\int_0^1 \log(1-t)\,\frac{(1-t)}{t}\,dt \qquad (x = te^{1-t}) \\ &= \frac{\pi^2}{6} - 1. \end{aligned}$$

In summary, we just proved that, as $z \to 1^-$

$$S(z, 1) = e^{L(z)} = a(1-z)^{\frac{1}{4}}\,\exp\left(\frac{c}{1-z}\right)(1 + o(1)) \quad (30)$$

where $a = \exp(-\frac{1}{4}\log\pi - \frac{1}{2}c)$.

So far, the main estimate (30) has been established as z tends to 1 from the left, by real values. In fact, (28) on which (30) rests holds for *complex* t only constrained in such a way that $-\frac{\pi}{2} + \epsilon \leq \arg(t) \leq \frac{\pi}{2} - \epsilon$, for any $\epsilon > 0$. The reason is that the converse mapping property (M4:ii) and, in particular, (28) rely on residues of the inverse Mellin integral that still converges when t is restricted to such a wedge (cf. [8]). Thus, the expansion (30) actually holds true as $z \to 1$ in a sector, say,

$$|\arg(1-z)| < \frac{\pi}{4}.$$

Saddle Point Analysis: It remains to collect the information gathered on $S(z, 1)$ and recover $s_n = [z^n]S(z, 1)$ asymptoti-

Input: A function $g(z)$ analytic in $|z| < R$ $(0 < R < +\infty)$ with nonnegative Taylor coefficients and "fast growth" as $z \to R^-$. Let $h(z) := \log g(z) - (n+1)\log z$.
Output: The asymptotic formula (B.2) for $g_n := [z^n]g(z)$ derived from the Cauchy coefficient integral

$$g_n = \frac{1}{2i\pi}\int_\gamma g(z)\,\frac{dz}{z^{n+1}} = \frac{1}{2i\pi}\int_\gamma e^{h(z)}\,dz \qquad (B.1)$$

where γ is a loop around $z = 0$.
(S1). SADDLE POINT CONTOUR. *Require that* $g'(z)/g(z) \to +\infty$ *as* $z \to R^-$. Let $r = r(n)$ be the unique positive root of the saddle point equation

$$h'(r) = 0 \qquad \text{or} \qquad \frac{rg'(r)}{g(r)} = n+1,$$

so that $r \to R$ as $n \to \infty$. The integral (B.1) is evaluated on $\gamma = \{z \mid |z| = r\}$.
(S2). BASIC SPLIT. *Require that* $h'''(r)^{1/3}h''(r)^{-1/2} \to 0$. Define $\phi = \phi(n)$ called the "range" of the saddle point by

$$\phi = \left|h'''(r)^{-1/6}h''(r)^{-1/4}\right|,$$

so that, as $\phi \to 0$, $h''(r)\phi^2 \to \infty$, and $h'''(r)\phi^3 \to 0$. Split $\gamma = \gamma_0 \cup \gamma_1$, where $\gamma_0 = \{z \in \gamma \mid |\arg(z)| \le \phi\}$, $\gamma_1 = \{z \in \gamma \mid |\arg(z)| \ge \phi\}$.
(S3) ELIMINATION OF TAILS. *Require that* $|g(re^{i\theta})| \le |g(re^{i\phi})|$ *on* γ_1. Then, the tail integral satisfies the trivial bound,

$$\left|\int_{\gamma_1} e^{h(z)}\,dz\right| = O\left(|e^{-h(re^{i\phi})}|\right).$$

(S4) LOCAL APPROXIMATION. *Require that* $h(re^{i\theta}) - h(r) - \frac{1}{2}r^2\theta^2h''(r) = O(|h'''(r)\phi^3|)$ *on* γ_0. Then, the central integral is asymptotic to a complete Gaussian integral, and

$$\frac{1}{2i\pi}\int_{\gamma_0} e^{h(z)}\,dz = \frac{g(r)r^{-n}}{\sqrt{2\pi h''(r)}}\left(1 + O(|h'''(r)\phi^3|)\right).$$

(S5) COLLECTION. Requirements $(S1), (S2), (S3), (S4)$, imply the estimate:

$$[z^n]g(z) = \frac{g(r)r^{-n}}{\sqrt{2\pi h''(r)}}\left(1 + O(|h'''(r)\phi^3|)\right) \sim \frac{g(r)r^{-n}}{\sqrt{2\pi h''(r)}}. \qquad (B.2)$$

Fig. 2. The saddle point algorithm.

cally. The inversion is provided by the Cauchy coefficient formula, that is,

$$s_n = \frac{1}{2\pi i}\oint \frac{S(z,1)}{z^{n+1}}\,dz$$

where the integration path is any simple loop around 0 inside the unit disk. The saddle point method [5], [11] summarized in Fig. 2 is now employed.

First, we provide a formula[1] for a standard set of functions that exhibit the same growth pattern as $S(z,1)$ near $z = 1$.

Lemma 4: For positive $A > 0$, and reals B and C, define $f(z) = f_{A,B,C}(z)$ as

$$f(z) = \exp\left(\frac{A}{1-z} + B\log\frac{1}{1-z} + C\log\left(\frac{1}{z}\log\frac{1}{1-z}\right)\right). \qquad (31)$$

Then, the nth Taylor coefficient of $f_{A,B,C}(z)$ satisfies asymptotically, for large n

$$[z^n]f_{A,B,C}(z) = \exp\Bigg[2\sqrt{An} + \frac{1}{2}\left(B - \frac{3}{2}\right)\log n + C\log\log\sqrt{\frac{n}{A}} - \frac{1}{2}\log\left(4\pi e^{-A}\big/\sqrt{A}\right)\Bigg](1 + o(1)). \qquad (32)$$

Proof: Problems of this kind have been considered by Wright [28] and others who, in particular, justify in detail that the saddle point method works in similar contexts. Therefore, we only outline the proof here. The starting point (see Fig. 2) is Cauchy's formula

$$[z^n]f(z) = \frac{1}{2\pi i}\oint e^{h(z)}\,dz$$

where

$$h(z) = \log f_{A,B,C}(z) - (n+1)\log z.$$

[1]The computations here and in the rest of the section have been further checked with the help of the symbolic system Maple. Note that this requires multiscale asymptotic manipulations for which the package based on the works of Salvy and Shackell [18] proved to be of special importance.

In accordance with (S1) of Fig. 2, one chooses a saddle-point contour that is a circle of radius r defined by $h'(r) = 0$. Asymptotically, one finds

$$r = 1 - \sqrt{\frac{A}{n}} + \frac{B-A}{2n} + o(n^{-1})$$

and

$$h(r) = 2A\sqrt{\frac{n}{A}} + B\log\left(\sqrt{\frac{n}{A}}\right) + C\log\log\left(\sqrt{\frac{n}{A}}\right) + \frac{1}{2}A + o(1).$$

The "range" $\phi = \phi(n)$ of the saddle point, where most of the contribution of the contour integral is concentrated asymptotically, is dictated by the order of growth of derivatives; see (S2). Here, $h''(r) \approx n^{3/2}$, while $h'''(r) \approx n^2$, so that

$$\phi(n) = n^{-3/4}.$$

In accordance with requirement (S3), tails are negligible since the function $\exp((1-z)^{-1})$ decays very fast when going away from the real axis. In the central region, the local approximation (S4) applies, as seen by expansions near $z = 1$. Thus, requirements (S1), (S2), (S3), and (S4) are satisfied, implying, by (S5)

$$[z^n]f(z) = \frac{1}{\sqrt{2\pi|h''(r)|}}\, e^{h(r)}\,(1 + o(1)).$$

Some simple algebra, using

$$h''(r) = 2n\sqrt{n/A}\,(1 + o(1))$$

yields the stated estimate (32). □

Now, the function $S(z, 1)$ is only known to behave like $f(z)$ of Lemma 4 in the vicinity of 1. In order to adapt the proof of Lemma 4 and legitimize the use of the resulting formula, we need to prove that $S(z, 1)$ decays fast away from the real axis.

Lemma 5 (Concentration Property): Consider the ratio

$$q(z) = \prod_{j=1}^{\infty} \left| \frac{\beta(z^j)}{\beta(|z|^j)} \right|.$$

Then, there exists a constant $c_0 > 0$ such that

$$q(re^{i\theta}) = O\left(e^{-c_0(1-r)^{-1}}\right)$$

uniformly, for $\frac{1}{2} \le r < 1$ and $|\arg(re^{i\theta} - 1)| > \frac{\pi}{4}$.

Proof: In this proof, the c_j denote positive constants whose precise value is immaterial.

First, by the triangular inequality, a function like $\beta(z)$ that has nonnegative Taylor coefficients attains its maximum modulus on the positive real axis. More precisely, one has

$$\sup_{\theta} \left|\beta(re^{i\theta})\right| = \beta(r).$$

Furthermore, by the converse triangular inequality, the maximum is uniquely attained on $|z| = r$ as soon as the function is aperiodic, which means the following. There is no $\hat{\beta}(z)$ analytic at 0 such that $\beta(z) = z^a\hat{\beta}(z^b)$ for integers a, b and $b \ge 2$. This condition is obviously satisfied here since $\beta(z) = 1 + e^{-1}z + 2(e^{-1}z)^2 + \cdots$.

Fix some small angle parameter ϕ_0, for instance, $\phi_0 = \frac{1}{10}$, and define

$$\sigma(r) = \sup_{|\theta| \ge \phi_0} \left| \frac{\beta(re^{i\theta})}{\beta(r)} \right|. \tag{33}$$

Then $\sigma(r)$ is continuous on the open interval $(0, 1)$ where it satisfies $\sigma(r) < 1$ while it tends to zero when r tends to 1. As a consequence, for each $\delta > 0$, there exists an $A_\delta < 1$ such that

$$\sigma(r) < A_\delta, \qquad \text{for all } r \text{ satisfying } \delta \le r < 1. \tag{34}$$

Consider the case where $z = re^{i\theta}$ with $r \to 1$. Set $r = e^{-\tau}$. The powers z^j form a discrete set of points on a logarithmic spiral that winds about 0. The number of such powers that have modulus larger than δ is

$$\frac{\log \delta^{-1}}{\tau} + O(1).$$

If $z = re^{i\theta}$ and $|\theta| \ge \phi_0$, then a fraction of these points, namely

$$c_1\,\frac{\log \delta^{-1}}{\tau} + O(1)$$

will lie outside of the region $|\arg(z^j)| < \phi_0$. Thus, by the bound (34), we find

$$q(re^{i\theta}) = O(e^{-c_2/\tau}) = O\left(e^{-c_3/(1-r)}\right). \tag{35}$$

This argument adapts when z is close to the real axis as follows. It is assumed that $|\arg(z-1)| > \pi/4$. Thus, $\arg(z) = \theta > \tau$. Then, the winding number around 0 of the polygonal line with vertices the z^j and $|z^j| > \delta$ is

$$\left\lceil \left(\frac{\log \delta^{-1}}{\tau} + O(1)\right)\frac{\theta}{2\pi} \right\rceil > \frac{\log \delta^{-1}}{2\pi} + O(1).$$

In other words, fixing δ small enough ensures that at least one full winding takes place. In this case, a number at least c_4/τ of the z^j satisfying $|z^j| > \delta$ are such that $|\beta(z^j)/\beta(r^j)| < A_\delta$. Then, an estimate of type (35) holds

$$\left|q(re^{i\theta})\right| = O(e^{-c_5/\tau}) = O\left(e^{-c_6/(1-r)}\right) \tag{36}$$

albeit with different constants. The statement follows upon taking $c_0 = \min(c_3, c_6)$. □

We are now finally ready to return to the estimate of s_n in Lemma 2. In the region $|\arg(z-1)| < \frac{\pi}{4}$, the Mellin asymptotic estimates (28) and (30) apply. This shows that in this region

$$S(z, 1) = e^{o(1)} f_{A,B,C}(z) \qquad (z \to 1),$$

where the function f is that of Lemma 4 and the constants A, B, C have the values assigned by (30)

$$A = c = \frac{\pi^2}{6} - 1 \qquad B = -\frac{1}{4} \qquad C = 0.$$

In the complementary region $|\arg(z-1)| > \frac{\pi}{4}$, the function $S(z, 1)$ is exponentially smaller than $S(|z|, 1)$ by Lemma 5. From these two facts, the saddle point estimates of Lemma 4 are seen to apply, by a trivial modification of the proof of that lemma. This concludes the proof of (18) in Lemma 2.

Moments: It remains to complete the evaluation of μ_n and σ_n^2, following the same principles as before. Start with $\mu_n = \mathbf{E}[K_n]$, with the goal of establishing the evaluation (19) of Lemma 2. It is necessary to estimate $[z^n]S'_u(z, 1)$, with

$$S'_u(z, 1) = S(z, 1)\sum_{k=0}^{\infty} z^k \frac{\beta'(z^k)}{\beta(z^k)}. \tag{37}$$

Let

$$D_1(z) = \sum_{k=0}^{\infty} \alpha(z^k), \qquad \text{where } \alpha(z) = z\,\frac{\beta'(z)}{\beta(z)}.$$

Via the substitution $z = e^{-t}$, the function $D_1(e^{-t})$ falls under the harmonic sum property (M3) of Fig. 1, so that its Mellin transform is

$$\mathcal{M}\left(D_1(e^{-t});\, s\right) = \zeta(s)\mathcal{M}\left(\alpha(e^{-t});\, s\right).$$

The asymptotic expansion

$$\alpha(e^{-t}) = \frac{1}{2t} - \frac{\sqrt{2}}{6}\,\frac{1}{\sqrt{t}} - \frac{1}{18} + O\left(\sqrt{t}\right)$$

gives the singular expansion of the corresponding Mellin transform, by (M4:i). This, in turn, yields the singular expansion of $\mathcal{M}(D_1(e^{-t});\, s)$. Then, the converse mapping property (M4:ii) gives back $D(e^{-t})$ at $t \sim 0$, hence,

$$D_1(z) = \frac{1}{2}\,\frac{1}{1-z}\,\log\frac{1}{1-z} + \frac{1}{2}\,\frac{\gamma}{1-z} - \frac{1}{6}\,\frac{\sqrt{2}\,\zeta(\frac{1}{2})}{\sqrt{1-z}} - \frac{1}{4}\log\frac{1}{1-z} + O(1)$$

where $\gamma = 0.577\cdots$ is the Euler constant. The combination of this last estimate and the main asymptotic form of $S(z,\,1)$ in (30) yields

$$S'_u(z,\,1) \underset{z\to 1}{\sim} \frac{1}{2}\,a\exp\left(\frac{c}{1-z} + \frac{3}{4}\log\frac{1}{1-z} + \log\log\frac{1}{1-z}\right) \tag{38}$$

where a is the same constant as in (30). Like for $S(z,\,1)$, the derivative $S'_u(z,\,1)$ is amenable to Lemma 4, and this proves the asymptotic form of μ_n, as stated in (19) of Lemma 2.

Finally, we need to justify (20) that represents a bound on the variance of K_n. The computations follow the same steps as above, so we only sketch them briefly. One needs to estimate a second derivative

$$\frac{S''_{uu}(z,\,1)}{S(z,\,1)} = D_2(z) + D_1^2(z)$$

where

$$D_2(z) = \sum_{k=0}^{\infty} z^{2k}\,\frac{\beta''(z^k)}{\beta(z^k)} - \left(\frac{z^k\beta'(z^k)}{\beta(z^k)}\right)^2.$$

The preceding sum is again a harmonic sum that is amenable to Mellin analysis, with the result that

$$D_2(z) = \frac{\zeta(2)}{2}\,\frac{1}{(1-z)^2} + O\left((1-z)^{-3/2}\right).$$

Then we appeal again to Lemma 4 to achieve the transfer to coefficients. Somewhat tedious calculations (that were assisted by the computer algebra system MAPLE) show that the leading term in $n\log^2 n$ of the second moment cancels with the square of the mean μ_n. Hence, the variance cannot be larger than $O(n\log n)$. This establishes the second moment estimate (20) of Lemma 2.

To complete the proof, we must show how to obtain the asymptotics for the upper bound on r_n^*, that is, $r_n^U = \sum_{i=0}^n r_i$. We follow in the footsteps of the analysis for r_n, that is, we define s_n^U, and observe that its generating function is

$$S^U(z,\,u) = \frac{S(z,\,u)}{1-z}.$$

In particular, (30) implies

$$S^U(z,\,1) = e^{L(z)} = a(1-z)^{-\frac{3}{4}}\exp\left(\frac{c}{1-z}\right)(1+o(1)).$$

Thus,

$$\log s_n^U = 2\sqrt{cn} - \frac{3}{8}\log n + O(1).$$

To establish $\log\mu_n^U$ (where μ_n^U is the corresponding quantity to μ_n defined in Lemma 2) we just repeat the calculations leading to (38) which yield

$$\log\mu_n^U = \frac{1}{2}\log n + \log\log n + O(1).$$

This establishes the desired upper bound and proves the main result.

V. Numerical Estimates

Numerical verifications support extremely well the claims made in the Introduction about the accuracy of asymptotic expansions based on analytic methods, even when the values of n are far from the asymptotic regime. Our main result states that the function

$$\varphi(n) = \frac{2}{\log 2}\sqrt{\left(\frac{\pi^2}{6}-1\right)n} - \frac{5}{8}\log_2 n + \frac{1}{2}\log_2\log n + K$$

where

$$K = \frac{1}{8}\lg\left(\frac{\pi^2/6-1}{2^{12}\pi^2}\right) \approx -1.99197$$

is such that $r_n = \varphi(n) + o(1)$. In fact, observation of the values of the difference $\Delta(n) = r_n - \varphi(n)$, for which a sample is given by the following table:

n:	3	5	10	20	50	100
$\Delta(n)$:	0.223	0.026	0.128	0.055	0.002	−0.010

shows that *the quantity* $\varphi(n)$ *estimates* r_n *very well for all values of* $n \geq 3$.

Acknowledgment

The authors wish to thank Y. Kontoyiannis (Purdue University) for pointing to them the paper of Csiszár and Shields. W. Szpankowski is also grateful to G. Seroussi and M. Weinberger (Hewlett-Packard Laboratories, Palo Alto, CA) for a very enjoyable and productive stay at Hewlett-Packard Laboratories at Palo Alto in summer 1997, where this work was initiated. The authors wish to thank B. Salvy (INRIA, France) for assistance with symbolic computations and are grateful to M. Drmota (University of Vienna, Austria) for many helpful comments. Finally, they express their deepest gratitude to an anonymous referee whose very careful reading allowed them to improve the paper. They also thank the Associate Editor Prof. I. Csiszár for his patience.

REFERENCES

[1] G. E. Andrews, *The Theory of Partitions*. Cambridge, U.K.: Cambridge Univ. Press, 1984.

[2] A. Barron, J. Rissanen, and B. Yu, "The minimum description length principle in coding and modeling," *IEEE Trans. Inform. Theory*, vol. 44, pp. 2743–2760, Oct. 1998.

[3] R. Corless, G. Gonnet, D. Hare, D. Jeffrey, and D. Knuth, "On the Lambert W function," *Adv. Comput. Math.*, vol. 5, pp. 329–359, 1996.

[4] I. Csiszár and P. Shields, "Redundancy rates for renewal and other processes," *IEEE Trans. Inform. Theory*, vol. 42, pp. 2065–2072, Nov. 1996.

[5] N. G. De Bruijn, *Asymptotic Methods in Analysis*. New York: Dover, 1981.

[6] M. Drmota and W. Szpankowski, "Generalized Shannon Code minimizes the maximal redundancy," in *Proc. LATIN'02 (Lecture Notes on Computer Science)*. Berlin, Germany: Springer-Verlag, 2002, vol. 2286, pp. 306–318.

[7] P. Flajolet and A. Odlyzko, "Singularity analysis of generating functions," *SIAM J. Discr. Meth.*, vol. 3, pp. 216–240, 1990.

[8] P. Flajolet, X. Gourdon, and P. Dumas, "Mellin transforms and asymptotics: Harmonic sums," *Theor. Comp. Sci.*, vol. 144, pp. 3–58, 1995.

[9] P. Flajolet and R. Sedgewick. Analytic Combinatorics book in preparation. [Online]. Available: http://pauillac.inria.fr/algo/flajolet/

[10] G. H. Hardy and S. Ramanujan, "Asymptotic formulae in combinatorial analysis," *Proc. London Math. Soc.*, ser. 2, vol. XVII, pp. 75–115, 1918.

[11] P. Henrici, *Applied and Computational Complex Analysis*. New York: Wiley, 1977, vol. 3.

[12] P. Jacquet and W. Szpankowski, "Asymptotic behavior of the Lempel–Ziv parsing scheme and digital search trees," *Theor. Comp. Sci.*, vol. 144, pp. 161–197, 1995.

[13] ——, "Entropy computations via analytic depoissonization," *IEEE Trans. Inform. Theory*, vol. 45, pp. 1072–1081, May 1999.

[14] P. Jacquet, W. Szpankowski, and J. Tang, "Average profile of the Lempel–Ziv parsing scheme for a Markovian source," *Algorithmica*, vol. 31, pp. 318–360, 2002.

[15] G. Louchard and W. Szpankowski, "On the average redundancy rate of the Lempel–Ziv code," *IEEE Trans. Inform. Theory*, vol. 43, pp. 2–8, Jan. 1997.

[16] A. Odlyzko, "Asymptotic enumeration," in *Handbook of Combinatorics*, R. Graham, M. Götschel, and L. Lovász, Eds. Amsterdam, The Netherlands: Elsevier, 1995, vol. II, pp. 1063–1229.

[17] J. Rissanen, "Complexity of strings in the class of Markov sources," *IEEE Trans. Inform. Theory*, vol. IT-32, pp. 526–532, July 1986.

[18] B. Salvy and J. Shackell, "Symbolic asymptotics: Multiseries of inverse functions," *J. Symb. Comput.*, vol. 27, pp. 543–563, 1999.

[19] S. Savari, "Redundancy of the Lempel–Ziv incremental parsing rule," *IEEE Trans. Inform. Theory*, vol. 43, pp. 9–21, Jan. 1997.

[20] P. Shields, "Universal redundancy rates do not exist," *IEEE Trans. Inform. Theory*, vol. 39, pp. 520–524, 1993.

[21] Y. Shtarkov, "Universal sequential coding of single messages," *Probl. Inform. Transm.*, vol. 23, pp. 175–186, 1987.

[22] W. Szpankowski, "On asymptotics of certain recurrences arising in universal coding," *Probl. Inform. Transm.*, vol. 34, no. 2, pp. 55–61, 1998.

[23] ——, *Average Case Analysis of Algorithms on Sequences*. New York: Wiley, 2001.

[24] ——, "Asymptotic redundancy of Huffman (and other) block codes," *IEEE Trans. Inform. Theory*, vol. 46, pp. 2434–2443, 2000.

[25] E. C. Titchmarsh, *The Theory of the Riemann Zeta Function*. Oxford, U.K.: Clarendon, 1951.

[26] Q. Xie and A. Barron, "Minimax redundancy for the class of memoryless sources," *IEEE Trans. Inform. Theory*, vol. 43, pp. 646–657, Mar. 1997.

[27] M. Weinberger, N. Merhav, and M. Feder, "Optimal sequential probability assignments for individual sequences," *IEEE Trans. Inform. Theory*, vol. 40, pp. 384–396, 1994.

[28] E. M. Wright, "The coefficients of certain power series," *J. London Math. Soc.*, vol. 7, pp. 256–262, 1932.

[29] A. J. Wyner, "The redundancy and distribution of the phrase lengths of the fixed-database Lempel–Ziv algorithm," *IEEE Trans. Inform. Theory*, vol. 43, pp. 1452–1464, Sept. 1997.

[30] J. Ziv, "Back from infinity: A constrained resources approach to information theory," *IEEE Inform. Theory Soc. Newslett.*, vol. 48, pp. 1, 30–33, Mar. 1998.